Including Linear Holding in Air Traffic Flow Management for Flexible Delay Handling

Yan Xu[a] and Xavier Prats[b]
Technical University of Catalonia, Castelldefels 08860, Barcelona, Spain

This paper introduces a strategy to include linear holding into air traffic flow management initiatives, together with the commonly used ground holding and airborne holding measures. In this way, flow management performance can be improved when handling delay assignment with uncertainty. Firstly, a trajectory generation method is presented, aiming at computing, per flight, the maximum linear holding realizable using the same fuel as the original nominal flight. This information is assumed to be computed and shared by the different airlines and it is then used to build a network air traffic flow management model to optimally assign delays, in the scope of trajectory based operations. Hence, the best distribution of delay is optimized at given positions along the flight trajectory (combining the three holding practices together) and taking into account the cost of delay, especially in the fuel consumption. The problem is formulated as a mixed integer linear program and solved with a commercial off-the-shelf solver. An illustrative example is given, showing that under the circumstance of capacity recovered ahead of schedule, including linear holding contributes to a notable delay reduction compared to the case where only ground and airborne holding apply.

[a] Ph.D. candidate, Department of Physics - Aeronautics Division, Office C3-121, Esteve Terradas 5, Castelldefels, Catalonia, Spain, Email address: yan.xu@upc.edu, AIAA Student Member.

[b] Associate Professor, Department of Physics - Aeronautics Division, Office C3-104, Esteve Terradas 5, Castelldefels, Catalonia, Spain, Email address: xavier.prats@upc.edu, AIAA Member.

Nomenclature

$f \in F$	set of flights
$t \in T$	set of time moments
$\tau \in \mathcal{T}$	set of periods for traffic demand
$T(\tau)$	τ^{th} time period defined within T
$k \in K$	set of airports
$w \in W$	set of sector entrance or exit positions
$j \in K \cup W$	set of positions
$s \in S$	set of sectors of capacity constrained
$P(f,i)$	$\begin{cases} \text{the departure airport,} & \text{if } i = 1 \\ \text{the arrival airport,} & \text{if } i = n_f \\ \text{sector positions,} & \text{if } 1 < i < n_f \end{cases}$
P_f	$\{P(f,i) : i \in [1, n_f]\}$, the positions of f along the scheduled trajectory in sequence
$R(f)$	$\{R(f,i) : i \in [1, n_f]\}$, the scheduled time of f in line with $P(f)$
r_f^j	$R(f,m)$, $m \in [1, n_f] : P(f,m) = j$, the scheduled time of f at j
T_f^j	$[r_f^j, r_f^j + e_f^j]$, the feasible time window for f at position j
e_f^j	the length of feasible time window
u^w	the maximum airborne holding time
$z_f^{j,j'}$	$r_f^{j'} - r_f^j$: $P(f,i) = j$, $P(f,i+1) = j'$, the scheduled duration of two contiguous positions
$v_f^{j,j'}$	the maximum LH bound of contiguous positions
$q_{f,f'}^k$	the minimum turnaround time
$D^k(\tau)$	the departure capacity of airport k in period τ
$A^k(\tau)$	the arrival capacity of airport k in period τ
$C^s(\tau)$	the capacity of sector s in period τ

I. Introduction

Air traffic flow management (ATFM) refers to processes of a more strategic nature, involving taking a higher-level view of the overall air traffic rather than controlling specific flights. It detects and resolves demand-capacity imbalances, smoothing aggregate traffic flows and keeping the workload of air traffic control (ATC) under manageable levels. For practical reasons, the institution

in charge of ATFM (known as the Network Manager in Europe) cannot take care of the specific preferences of one particular flight, since the overall objective of ATFM is typically to reach a global optimum (e.g., minimize total delay across all controlled flights) based on some unanimous fairness criteria (e.g., first scheduled, first served).

With the paradigm shift for the future air traffic management (ATM) proposed by SESAR (Single European Sky ATM Research) in Europe and NextGen (Next Generation Air Transportation System) in the United States, with a transition from airspace based operations to trajectory based operations (TBO), the airspace users will be expected to increasingly participate in ATM decisions using, in particular, more collaborative decision making (CDM) mechanisms. For example, the SESAR concept of reference business trajectory (RBT), as output of an ATFM negotiation, is the trajectory that the airspace user agrees to fly and the ANSP (air navigation service provider) and airports agree to facilitate. In addition, as one current effort toward generating and sharing alternative trajectories to aid in CDM operations, the Collaborative Trajectory Options Program (CTOP) developed by the FAA (Federal Aviation Administration) and airlines has completed its testing and is being deployed, which has been aimed at balancing en route traffic demand with available capacity in the National Airspace System (NAS) [1, 2].

Under current Ground Delay Programs (GDPs), one of the most sophisticated ATFM tools used in the United States, resources (i.e., arrival slots) are assigned to flights in accordance with a ration-by-schedule (RBS) mechanism (first-scheduled first-served prioritization). It is accompanied with CDM initiatives, such as flight substitution, cancellations, compression, or slot credit substitution, allowing airlines to manage their own flights in line with their specified policies [3]. Following this thought, we could imagine that airlines would be willing to provide specific flight information to the Network Manager, especially if potential benefits might exist (e.g., reducing flight delays) along with feasible negotiation mechanisms (e.g., the TBO paradigm).

In this paper, we discuss the applicability of including a cost-based aircraft speed control, or linear holding (LH), as specific information provided by airlines to the Network Manager, in the ATFM delay assignment process and aiming at increasing the flexibility when handling delays and improving the performance of ATFM regulations. The LH strategy was proposed in [4], aimed at

partially incurring in the air (by flying slower) the initially assigned ground delays. This strategy was further explored in [5], where aircraft were allowed to cruise at the lowest possible speed at which the fuel consumption remained exactly the same as initially planned. In this situation, if the delays were canceled ahead of schedule, aircraft already airborne and flying slower, could speed up to the initially planned speed and recover part of the delay without extra fuel costs [6]. Recently, an aircraft trajectory optimization technique was adopted in [7], where the whole flight profile was subject of optimization to maximize the achievable LH at no extra fuel cost.

As the core method to perform LH, the speed adjustment has proven successful for several ATM scenarios. For instance, a speed control approach was presented in [8] to transfer delay away from the terminal to the en route, from which significant fuel saving on a per flight basis was also yielded. In [9], a pre-tactical speed control was applied en route to prevent aircraft from performing airborne holding when arriving at a congested airspace. Similarly, but more at the tactical level, aircraft in [10] were required to reduce their speed to avoid arriving at the airport before its opening time to reduce unnecessary holdings.

On the other hand, following the pioneering work done in [11], a number of researchers have focused their activity on the development of optimization models for the delay assignment as a short-term measure for traffic flow regulation (see [12] for instance). Further taking into account the capacity constraints from airspace sectors, the problem of controlling release times and speed adjustments of aircraft while airborne for a network of airports (including sectors) has been studied in [13–15]. Meanwhile, the optimal control of traffic flow in the terminal area has been targeted by research and development efforts as well, particularly in terms of runway scheduling, aircraft (re)routing and taxiing planning, to reduce flight delays, minimize energy consumptions and mitigate airport congestions [16–19].

The network ATFM model in this paper is based on the above mentioned Bertsimas and Stock-Patterson model [13], which has been widely studied in the last decades. Up to the best of our knowledge, however, among past research less discussion on the operating cost impact of such speed adjustments as delay absorption measure has been done. Given the fact that aircraft speeds are intimately related to fuel consumption, the main contribution of this paper is to include a cost-based

LH in this well-known model, and exploit its flexibility to improve ATFM performance.

II. Linear holding practices

In this section the linear holding concept is explained and a methodology, based on trajectory optimization algorithms, is outlined with the purpose to compute the maximum linear holding time without incurring extra fuel consumption.

A. Linear holding concept

Fig. 1 highlights schematically the main characteristics of linear holding together with the two commonly used holding practices in current ATM: ground and airborne holding. Typical airborne holding would consume more fuel due to the extended flight track (assuming no specific speed adjustment), whilst ground holding has no impact in fuel consumption. Due to the increased extra fuel, the airborne holding time is fairly limited, taking account that safety related issues may arise from a reduction of the on-board reserve fuel.

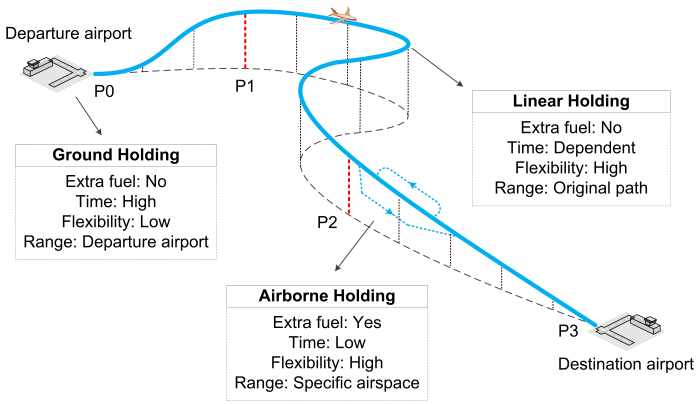
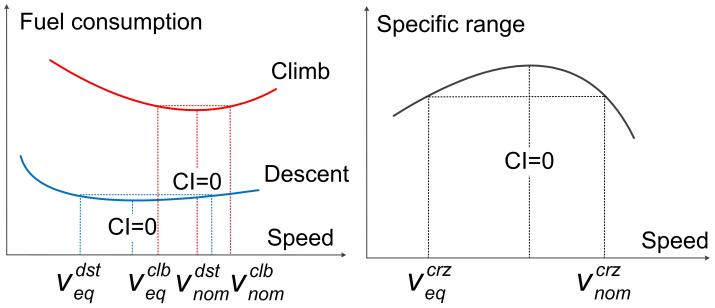


Fig. 1 A comparison between ground, airborne and linear holding.

Airlines consider direct operating costs (DOC) when planning their flights [20], which besides fuel consumption also take into account time-related costs. In this context, on-board flight management systems (FMS) allow them to optimize trajectories by means of a Cost Index (CI) input parameter, which expresses the ratio between time-related costs and the cost of fuel. Flying at a CI greater than zero (typically the case) will result in a speed greater than the maximum range speed, since time savings will be also considered.

Notionally, an equivalent airspeed (V_{eq}) can be defined by that speed, lower than the speed

initially planned for the nominal flight (V_{nom}), which produces the same fuel consumption as flying at V_{nom} (see Fig. 2). Thus, by flying between these two speeds the fuel burnt will not exceed the quantity initially planned, while some LH could be realized. It should be noted that the (reduced) equivalent airspeed might be overridden by the lowest selectable speed V_{LS}, namely the stalling speed at 1.3g, in autopilot. However, as indicated in previous work [7], a more stable Green Dot (GD) airspeed is suggested as the lower bound, which depicts the best lift to drag ratio speed in clean configuration and that is higher than the V_{LS} [21]. Generally, the amount of LH that can be realized depends on several factors, such as the aircraft type, trip distance, payload, cruise flight level, etc. This topic has been discussed in detail by [7], exploring the maximum achievable LH for some particular flights on the previous premise of no extra fuel consumption.



(a) Climb (clb) and descent (dst) (b) Cruise (crz)

Fig. 2 Definition of the equivalent speed (V_{eq}).

From an implementation point of view, ground holding can only be performed at the departure airport, prior to take-off. Airborne holding (including holding patterns or path stretching) can be done at any available airspace, in theory, but practically it is typically performed in designated locations. The most promising feature of LH is that delay absorption can be flexibly managed through proper speed adjustment along the original route, without incurring extra fuel burns than initially scheduled if flying between V_{eq} and V_{nom}. Note that LH could also be performed by burning more fuel, as shown in [5]. Yet, it is out of the scope of this paper to explore the benefits of this strategy. Moreover, airline operators should always compare if the advantages of LH can compensate for the possible drawbacks of flying at lower speeds (e.g., more engine maintenance might be required) based on their specific situations, and eventually decide whether and how they

will participate in the CDM process (which will be discussed in Sec. III B).

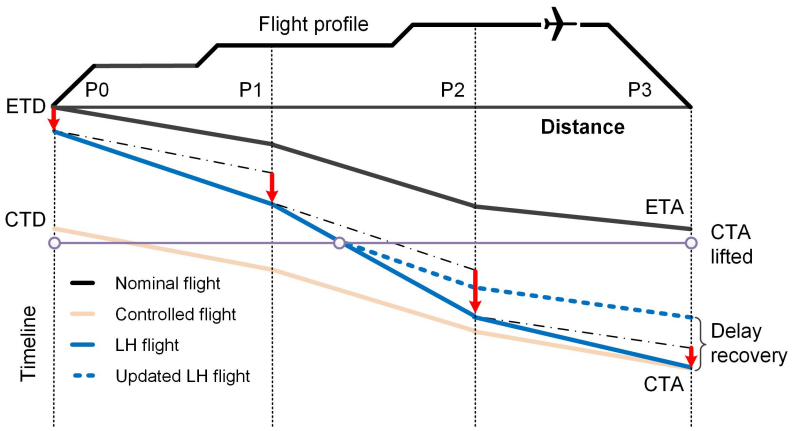


Fig. 3 Schematic of a potential applicability of LH for ATFM.

To see the potential applicability of LH, let us imagine a flight assigned with a certain delay as a result of a GDP, as shown in Fig. 3. In the near future, we could assume that controlled times of arrival (CTA) could be enforced at the destination airport, in order to guarantee the arrival slot allocation computed by the GDP [22]. Our flight could absorb all delay by means of ground holding, as currently done (orange line in Fig. 3); or perform less ground holding but some LH in the air, such that the CTA is still met (blue solid line). If we assume that at some point the GDP is canceled, due to weather improvement for instance (purple line), airborne aircraft could stop the LH, accelerate to the nominal speed, and recover part of its delay (blue dash line). Obviously, if the CTA is not changed, aircraft will finally arrive at the destination airport with the same amount of delay as in the case where all delay is served on the ground.

Under the circumstance that more (every) aircraft fly slower (to perform LH) absorbing part of the delay that could have been realized by ground holding, the airborne traffic density will accordingly increase. This may in the long run cause heavier workloads to air traffic controllers who are responsible to guarantee aircraft separation. However, from the ATFM point of view, higher airborne density could instead contribute to better utilization of airspace sector capacities, and is what we expect to achieve in this paper by implementing LH. It is worth noting that the traffic density's growth does not necessarily increase traffic throughput since the latter is equal to density multiplied by flow speed (which has been reduced in this case).

B. Trajectory optimization for linear holding computation

The optimization of an aircraft trajectory requires the definition of a mathematical model representing aircraft dynamics, along with a model for certain atmospheric parameters. In this paper, a point-mass aircraft model, an enhanced performance model using manufacturer certified data, and the International Standard Atmosphere (ISA) have been considered. For more details, the readers may refer to [23].

A generic vertical trajectory can be partitioned into several segments $i \in \{1, \cdots, N\}$, where different constraints or models may apply. Fig. 4 shows the different segments considered in this paper, where the initial and final points are taken, respectively, at the moment the slats are retracted (after take-off) and extended (before landing).

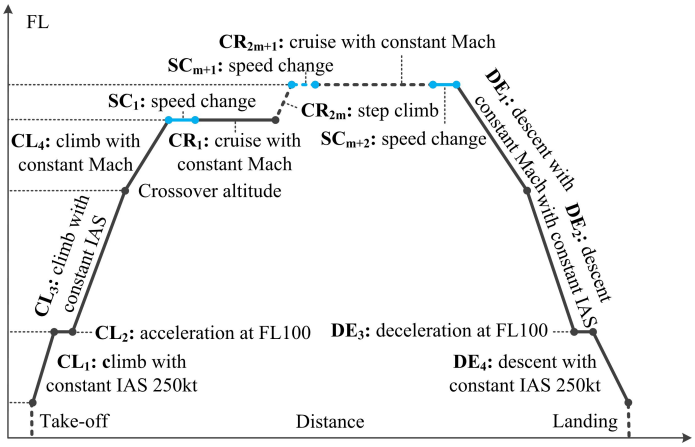


Fig. 4 Model for the vertical profile used in the trajectory optimization tool.

Before considering LH, the nominal trajectory (i.e., that one initially scheduled by the airline) is generated first. In line with the discussion in Sec. II A, the optimal trajectory will be that one minimizing a compound cost function J over the whole time window $[t_0^1, t_f^N]$, as follows:

$$\min J = \min \int_{t_0^{(1)}}^{t_f^{(N)}} (FF(t) + CI)\mathrm{d}t, \tag{1}$$

where $FF(t)$ is the fuel flow and CI is the Cost Index, combined to reflect the airline's DOC (i.e., time related costs and the cost of fuel consumption).

This optimization problem is subject to several constraints, including the aircraft dynamics,

in the form of differential equations on the aircraft states; some algebraic event constraints, fixing initial/final states; path constraints, in order to fulfill ATM and operational constraints along the different flight phases (see Fig. 4); and finally, link constraints, to ensure continuity along the flight segments. Then, a non-linear programming (NLP) problem can be formulated and solved with a commercial off-the-shelf solver. A detailed description of all these constraints and a complete formulation of the problem is given in [23].

Then, based on the nominal trajectory, by changing the optimization objective and constraints, the trajectory can be modified performing some LH. As an example, we will illustrate the case where the maximum LH is realized at no extra fuel cost. The objective function of the problem (1) is replaced by (2), where the total flight time is maximized, whilst subject to an additional constraint on the total fuel consumption namely the fuel consumption being less than that of the nominal trajectory, as depicted by (3):

$$\max J = \max \int_{t_0^{(1)}}^{t_f^{(N)}} \mathrm{d}t, \tag{2}$$

$$s.t. \int_{t_0^{(1)}}^{t_f^{(N)}} FF(t)\mathrm{d}t \leq F_{nom}, \tag{3}$$

where F_{nom} is the fuel consumed in the nominal trajectory.

This makes it clear that the flight, as a whole, is optimized rather than the climb, cruise or descent phases separately (or even the subdivided flight segments within a particular phase). In this paper, however, the aircraft trajectory is allowed to update tactically (i.e., partway en route) in response to possible changes of ATFM regulations, such as the cancellation of controlled times of arrival due to prior weather clearance, as shown in Fig. 3. In order to keep fuel consumption similar to the nominal flight, however, the mass of the aircraft M is fixed for each trajectory segment i defined in the vertical flight profile of Fig. 4:

$$M_{LH}^{(i)} = M_{nom}^{(i)}, i = CL_1, ..., DE_4. \tag{4}$$

Otherwise, for instance, more fuel might be burnt during climb (in the LH trajectory), leaving less fuel available for updating the remaining flight phases (when the situation may have improved).

In addition, the cruise flight level(s) and route should be fixed as well, as pre-tactical re-routings or flight level cappings, as part of a possible ATFM negotiation, are out of the scope of this paper:

$$H_{LH}^{CL_4} = H_{nom}^{CL_4}, H_{LH}^{DE_1} = H_{nom}^{DE_1}, H_{LH}^{CR_{2m}} = H_{nom}^{CR_{2m}}, \tag{5}$$

where H denote the aircraft flight altitude. Since the angle of climb (descent) varies with speed, the climb (descent) distances will be different at different speeds, meaning that the location of the top of climb (TOC) and top of descent (TOD) will also move. Moreover, the distance at which each step climb (if any) is performed will not be enforced, considering that possible changes in the TOC and/or TOD could impact on the length of different cruise segments.

Finally, it should be noted that before each cruise flight level, a short cruise segment (with a maximum duration of 1 minute) is added in order to allow speed adjustments (see Fig. 4). A similar segment is added at the end of the last cruise phase in order to adjust to the optimal descent Mach. These segments help to reduce the excessive influences from the link constraints on the flight profile.

III. Network ATFM model with linear holding

A network ATFM model is proposed in this section, which assigns delays at designed positions. Ground and airborne holdings remain the default measure for delay absorption, while linear holding is possible for those airlines willing to participate in the ATFM delay assignment, by generating and sharing certain information with the Network Manager to aid in an outlined CDM process.

A. Problem statement

Under existing ATFM network models (see [13] and the references therein), delays could be assigned to flights by means of ground or airborne holding. Airborne holding, however, tends to be less preferred (especially when a long-time delay occurs) because of its higher fuel costs (and potential safety issues) as has been discussed in Sec. II A.

In the model proposed in this paper, we maintain the above two holding practices, but add the LH option. Nevertheless, aiming at differentiating the proposed LH with typical airborne holding,

the delay assignment is conducted at specific designed "positions" along the scheduled trajectory, by using the concept of Controlled Time of Arrival (CTA) and Controlled Time of Departure (CTD) at each position. Accordingly, the decision variables of the model are defined as follows:

$$x_{f,t}^{j} = \begin{cases} 1, & \text{if flight } f \text{ departs from the position } j \text{ by time } t \\ 0, & \text{otherwise} \end{cases}$$

$$y_{f,t}^{j} = \begin{cases} 1, & \text{if flight } f \text{ arrives at the position } j \text{ by time } t \\ 0, & \text{otherwise} \end{cases}$$

Note that the "by" time is used, rather than "at" as the decision variables in this paper, which would enable a faster solution searching time according to [13], while the "at" time can be derived by $(x_{f,t}^{j} - x_{f,t-1}^{j})$ and $(y_{f,t}^{j} - y_{f,t-1}^{j})$ respectively. To enforce that only one time slot will be assigned to one flight at each designed position, within a prescribed feasible window T_f^j, it has to satisfy $\sum_{t \in T_f^j} (x_{f,t}^{j} - x_{f,t-1}^{j}) = 1$ and $\sum_{t \in T_f^j} (y_{f,t}^{j} - y_{f,t-1}^{j}) = 1$. However, when using the "by" time, this constraint can be simplified as to $x_{f,\overline{T}_f^j} = 1$, $y_{f,\overline{T}_f^j} = 1$, and $x_{f,\underline{T}_f^j - 1} = 0$, $y_{f,\underline{T}_f^j - 1} = 0$, where $\underline{T}_f^j$ and $\overline{T}_f^j$ are respectively the lower and upper bound of the feasible solution, namely $[\underline{T}_f^j, \overline{T}_f^j] = T_f^j$.

Fig. 5 shows schematically flight time versus distance and the three types of holding strategies: ground holding is performed only at the origin airport; airborne holding can only be performed "at" a given position (the difference between the "departure" and "arrival" time at that position equals to the holding time); and since LH is performed by flying slower, the slope of the lines is increased if compared with the planned schedule.

Recall that we distinguish the typical airborne holding from LH by the fact that when performing the former, the actual flight distance will be extended (either by vectoring or using holding patterns). This flight path "stretching", however, does not contribute to the execution of trajectory defined by each contiguous point. Thus, the typical airborne holding, on some level, can be seen as a "circling" at a particular position.

It is also worth noting that the "positions" referred here (such as P1, P2 and P3 in Fig. 5) may not correspond to the actual geographical waypoints existing in current airspace. The model

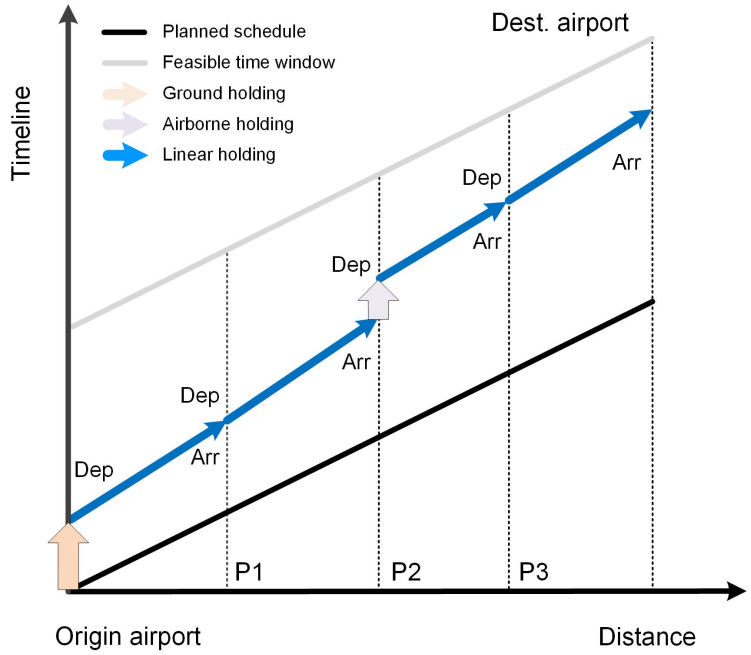


Fig. 5 Characteristics of ground, airborne and linear holding in the ATFM network model proposed in terms of flight time versus distance.

in this paper defines entrance and exit positions at each elementary sector that the controlled flight is scheduled to traverse (as well as the two representing origin and destination airports respectively, as shown in Fig. 6), in such a way that the traffic demand of each sector and airport (for departure and arrival) during different time periods can be managed under capacity constraints. In addition, the feasible time window shown in Fig. 5 defines a solution space based on the flight schedule, which will largely reduce the number of variables taken into optimization, and that in turn can be further discretized to prescribed time steps.

B. Participation of airlines in the ATFM process

As indicated in Sec. II B, the amount of delay absorption that LH can realize is constrained by the fuel consumption, which again is dependent on the aircraft type, take-off mass, flight distance, etc. Thus, from the ATFM perspective, considering all these data would be a daunting task. Moreover, some of the airline's information is proprietary, such as aircraft mass and fuel consumption figures, which is normally not publicly accessible. From the airline perspective, however, they could have a clear view of all the information of their own flights, and thus have an intimate knowledge

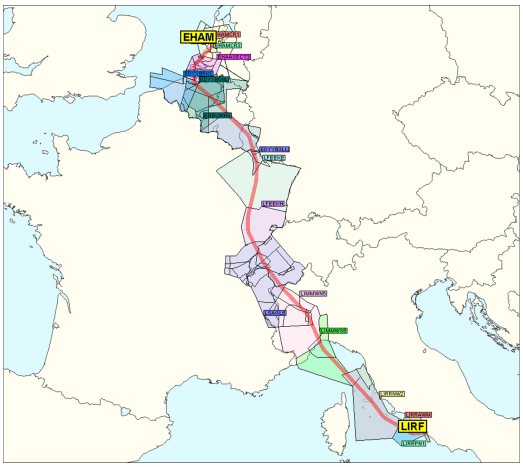


Fig. 6 An example of a scheduled 3D trajectory from LIRF to EHAM airports, which traverses multiple contiguous sectors. (Source: Eurocontrol's NEST modelling tool)

of the capability of each particular flight to absorb delays airborne.

Table 1 Flight route extracted from current planning information.

Enter Time	Airport	Ele. Sector	Level	Crossed Dur.	Exit Time
-	LIRF	-	0	0:00:00	7:59:00
8:00:36	-	LIRRPN1	35	0:06:42	8:07:18
8:08:10	-	LIRRNW2	208	0:06:08	8:14:18
8:14:18	-	LIRRNW3	284	0:01:50	8:16:08
8:16:08	-	LIRRNW4	304	0:03:01	8:19:09
8:19:09	-	LIRRNW5	334	0:05:48	8:24:57
8:24:57	-	LIMMWS5	360	0:08:47	8:33:44
8:33:44	-	LIMMWN5	360	0:14:14	8:47:58
8:47:58	-	LSAGL5	360	0:11:31	8:59:29
8:59:29	-	LFEEKH	360	0:07:21	9:06:50
9:06:50	-	LFEEKE	360	0:07:54	9:14:44
9:14:44	-	EDYYB3LH	360	0:09:03	9:23:47
9:23:47	-	EDYYB3LL	335	0:05:40	9:29:27
9:29:27	-	EDYYB5NL	300	0:08:41	9:38:08
9:39:47	-	EHAASECT3	230	0:07:23	9:47:10
9:47:10	-	EHAMTA	85	0:04:26	9:51:36
9:52:18	-	EHAMCR1	28	0:03:22	9:55:40
9:55:40	EHAM	-	0	0:00:00	-

Consequently, aiming at including the proposed LH into delay assignment, the ATFM model in this paper requires (only) one more input from airlines than those models found in the literature: the maximum LH bound per flight, along the planned trajectory. Given the procedure in practice, as a matter of fact, the Network Managers currently offer a high level of flexibility to airlines for flight planning [24]. In addition, in case that no data for this input is provided, the model could

still work by setting the default value for LH to zero.

Let us first take a look at what could be found from current flight planning information, as provided by the demand data repository v2 (DDR2) published by EUROCONTROL. Table 1 presents the detailed information obtained for a specific flight scheduled from Rome Fiumicino Airport (LIRF) to Amsterdam Schiphol Airport (EHAM) shown in Fig. 6. These data include "enter time", "exit time", "crossed duration" and "altitude" at each of the sectors (and airports) the aircraft is scheduled to fly, which correspond to the designed "positions" (sector boundaries and airports) used in our model.

Based on the planned vertical profile found in DDR2, the nominal optimal trajectory has been reconstructed, on one hand, using the trajectory optimization methodology described in Sec. II B (red line in Fig. 7). On the other hand, the blue line in the figure represents the same flight when performing the maximum amount of LH, while incurring the same fuel consumption, during every single flight segment (i.e., each discretized flight phase, as shown in Fig. 4, in accordance with typical ATM regulations) as enforced by Eq. 4 in the trajectory optimization. Fig. 7 also shows the true airspeed (TAS) of both trajectories, where the airborne delay generation by means of LH can be easily seen.

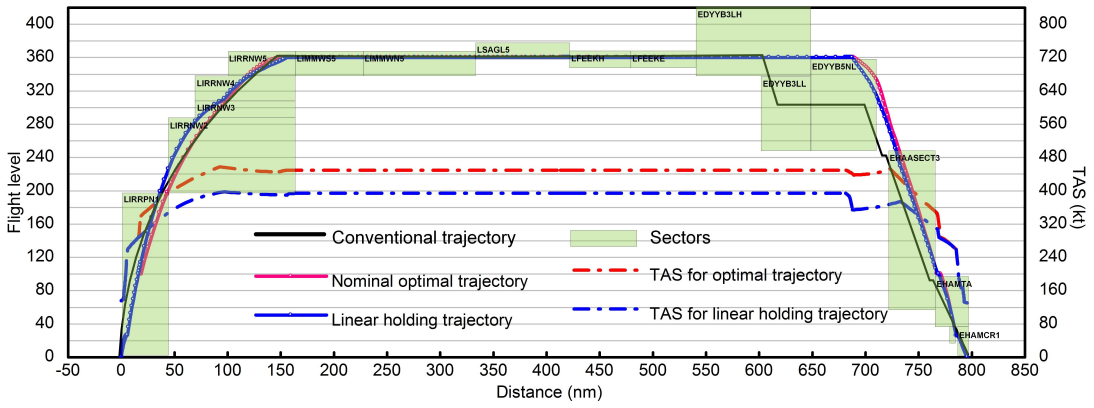


Fig. 7 Vertical and speed profiles of the nominal and LH trajectories traversing the scheduled contiguous sectors.

Table 2 summarizes the trajectories of Fig. 7 in form of the current flight planning information (as Table 1), but is added with a "LH time" bound (see the rightmost column), which equals to

the difference of crossed duration between the nominal trajectory and the LH trajectory (negative values appear in climb/descent because of the slight differences on the trajectory caused from speed changes), and that is the one that should be provided (by airlines) to the Network Manager. It is worth noting that, the crossed segment shown in Table 2 represents the distance flown between the entry and exit of a particular sector, which differs from the flight segment mentioned above incurring the same fuel between the nominal and LH trajectories, and thus there appear some slight differences in fuel. Still, they should be similar and at the end of the trajectory always be the same with respect to the total fuel burns.

Table 2 Nominal and LH trajectories flight planning information.

Position	Nominal optimal trajectory							Linear holding trajectory							
	Enter Position		Exit Position		Crossed Segment			Enter Position		Exit Position		Crossed Segment			
	Dist (nm)	Time (hh:mm:ss)	Dist (nm)	Time (hh:mm:ss)	Length (nm)	Duration (mm:ss)	Fuel (kg)	Dist (nm)	Time (hh:mm:ss)	Dist (nm)	Time (hh:mm:ss)	Length (nm)	Duration (mm:ss)	Fuel (kg)	*LH time (min)*
LIRF	-	-	0	7:59:00	-	-	-	-	-	0	7:59:00	-	-	-	**-**
LIRRPN1	2	7:59:50	46	8:07:53	44	08:03	778	2	7:59:50	44	8:08:01	42	08:11	760	***0.13***
LIRRNW2	48	8:08:28	80	8:13:06	32	04:38	333	48	8:08:42	70	8:12:53	22	04:11	258	***-0.45***
LIRRNW3	80	8:13:06	98	8:15:14	18	02:08	136	70	8:12:53	87	8:14:59	17	02:06	105	***-0.03***
LIRRNW4	98	8:15:14	115	8:17:34	17	02:20	135	87	8:14:59	115	8:19:26	28	04:27	208	***2.12***
LIRRNW5	115	8:17:34	165	8:24:26	50	06:52	311	115	8:19:26	165	8:27:03	50	07:37	311	***0.75***
LIMMWS5	165	8:24:26	229	8:32:43	64	08:17	322	165	8:27:03	229	8:37:45	64	10:42	322	***2.42***
LIMMWN5	229	8:32:43	334	8:46:54	105	14:11	549	229	8:37:45	334	8:56:05	105	18:20	549	***4.15***
LSAGL5	334	8:46:54	423	8:58:33	89	11:39	447	334	8:56:05	423	9:12:21	89	16:16	493	***4.62***
LFEEKH	423	8:58:33	479	9:05:46	56	07:13	276	423	9:12:21	479	9:21:29	56	09:08	276	***1.92***
LFEEKE	479	9:05:46	538	9:14:10	59	08:24	320	479	9:21:29	538	9:30:37	59	09:08	275	***0.73***
EDYYB3LH	538	9:14:10	649	9:28:36	111	14:26	546	538	9:30:37	649	9:50:23	111	19:46	567	***5.33***
EDYYB5NL	692	9:34:38	708	9:36:49	16	02:11	59	684	9:56:48	708	10:00:39	24	03:51	71	***1.67***
EHAASECT3	729	9:39:42	763	9:44:41	34	04:59	34	727	10:03:50	763	10:10:13	36	06:23	46	***1.40***
EHAMTA	774	9:46:25	786	9:51:44	12	05:19	35	773	10:11:26	786	10:17:09	13	05:43	39	***0.40***
EHAMCR1	786	9:51:44	796	9:55:28	10	03:44	85	786	10:17:09	796	10:20:52	10	03:43	85	***-0.02***
EHAM	796	9:55:28	-	-	-	-	4366	796	10:20:52	-	-	-	-	4366	***25.13***

C. Model formulation

The network ATFM model with the above presented airline-enabled LH is formulated in the following section. As mentioned before, the overall framework of the model is based on the widely-studied Bertsimas and Stock-Patterson model [13].

1. Objective function

In this model, the cost of the total delay (TD) is minimized including the costs consequence of ground holding (GH), airborne holding (AH) and linear holding (LH):

$$\min(cost_{TD}) = \min(GH + \alpha AH + \beta LH), \tag{6}$$

where α and β are the cost weighting factors. Since $TD = GH + AH + LH$, we can substitute LH

in (6), yielding to:

$$\min(cost_{TD}) = \min[\beta TD + (\alpha - \beta)AH + (1 - \beta)GH)]. \tag{7}$$

Taking into account the fairness of delay assignment, as discussed in [15], the total delay is multiplied by a coefficient $c_f = (t - r_f^k)^{1+\epsilon}, \epsilon > 0$ in (7). In this way, delays will be assigned moderately across all the flights, instead of unevenly to one particular flight. Accordingly, the objective function can be arranged as:

$$\min \sum_{f \in F} [\beta c_f h_f + (\alpha - \beta)a_f + (1 - \beta)g_f],$$

$$c_f h_f = \sum_{t \in T_f^k, P(f,n_f)=k} (t - r_f^k)^{1+\epsilon} \cdot (y_{f,t}^k - y_{f,t-1}^k),$$

$$a_f = \sum_{t \in T_f^w, w \in P(f,i): 1<i<n_f} t \cdot (x_{f,t}^w - x_{f,t-1}^w - y_{f,t}^w + y_{f,t-1}^w),$$

$$g_f = \sum_{t \in T_f^k, P(f,1)=k} (t - r_f^k) \cdot (x_{f,t}^k - x_{f,t-1}^k). \tag{8}$$

The constraints of this model can be grouped into flight operations, network capacities, decision variables and delay updates, as presented in each subsection below. It is worth noting that, for updating the delay (assignment), different from state-of-the-art stochastic dynamic models (see for instance in [25]), full deterministic information (e.g., weather forecast) is assumed in this paper, such that it is feasible to realize the dynamic updating by re-executing the model (by means of further including specific constraints, i.e., Constraints 19, 20, 21 and 22).

2. Flight operations constraints

$$x_{f,t}^j - x_{f,t-1}^j \geq 0 \quad \forall f \in F, \forall j \in P_f, \forall t \in T_f^j, \tag{9}$$

$$y^j_{f,t} - y^j_{f,t-1} \geq 0 \quad \forall f \in F, \forall j \in P_f, \forall t \in T^j_f, \tag{10}$$

$$x^j_{f,t} - y^j_{f,t} \leq 0 \quad \forall f \in F, \forall w \in W, \forall t \in T^w_f, \tag{11}$$

$$\begin{aligned} y^{j'}_{f,t'} - x^j_{f,t} \leq 0 \quad & \forall f \in F, \forall i \in [1, n_f - 1], P(f,i) = j, P(f,i+1) = j', \\ & \forall t \in T^j_f, t' = t + z^{j,j'}_f, \end{aligned} \tag{12}$$

$$\begin{aligned} y^{j'}_{f,t'} - x^j_{f,t} \geq 0 \quad & \forall f \in F, \forall i \in [1, n_f - 1], P(f,i) = j, P(f,i+1) = j', \\ & \forall t \in T^j_f, \forall t' \in T^{j'}_f, t' = t + z^{j,j'}_f + v^{j,j'}_f, \end{aligned} \tag{13}$$

Constraints (9) and (10) ensure that each flight f is assigned with only one slot for departing and arriving, respectively, at position j. Constraint (11) imposes a maximum airborne holding time u^w at each designed position. Constraint (12) enforces that LH to be non-negative (i.e., flying faster than initially planned is not considered for delay assignment in this model). This is because, as discussed in Sec. II B, the on-board flight management system could help airlines to optimize the aircraft trajectories by setting the CI input, which reflects airlines preferences (or trade-offs) on speed and fuel burns when planning their flights, and thus these initially scheduled speeds, should already be the highest that are favored by airlines. Constraint (13) stipulate that the LH performed between two contiguous positions of flight f should not exceed the maximum LH bound $v^{j,j'}_f$, which is provided by airlines and that is set by 0 as default if such information is not provided.

3. Network capacity constraints

$$\sum_{f \in F: P(f,1) = k} \sum_{t \in T^k_f \cap T(\tau)} (x^k_{f,t} - x^k_{f,t-1}) \leq D^k(\tau) \quad \forall k \in K, \forall \tau \in \mathcal{T}, \tag{14}$$

$$\sum_{f \in F: P(f,n_f) = k} \sum_{t \in T^k_f \cap T(\tau)} (y^k_{f,t} - y^k_{f,t-1}) \leq A^k(\tau) \quad \forall k \in K, \forall \tau \in \mathcal{T}, \tag{15}$$

$$\sum_{f\in F:P(f,i)=w,i\in[1,n_f-1]}\sum_{t\in T_f^w\cap T(\tau)}(x_{f,t}^w-x_{f,t-1}^w)\leq C^s(\tau)\quad\forall w\in s\subset S,\forall\tau\in\mathcal{T}, \tag{16}$$

Constraints (14), (15) and (16) ensure that the traffic demand would not exceed the capacity of departure airport, arrival airport and en route sectors, respectively. It is worth noting that the flight performing airborne holding in this model is counted within the boundary of its current sector (i.e., before departing the position). Since the capacity values are all defined within a period of time window, they are capable of being modified following the changes of the network environment, such as the improvement of weather conditions or traffic situations.

4. Constraints on decision variables

$$x_{f,t}^j\in 0,1\quad\forall f\in F,\forall j\in P_f,\forall t\in T_f^j, \tag{17}$$

$$y_{f,t}^j\in 0,1\quad\forall f\in F,\forall j\in P_f,\forall t\in T_f^j. \tag{18}$$

Constraints (17) and (18) state that the decision variables of the model are binary.

Above all, the model can be modified to perform the iterative delay assignment. Assume at the start of the $(\tau+1)$ th time period, i.e., t_σ, the capacity changes from current status of the time period of $T(\tau')$, and requires for another round of delay assignment. We could simply fix part of the decision variables based on the current results, and optimize the rest of them in the next round of delay assignment. However, as previously mentioned, full deterministic information is assumed in this study with respect to capacity updates, for specifically tactical ATFM scenario, which may cause the model to run the risk of failure as the uncertainty grows if applied to a strategic stage.

5. Constraints from updating assignment

$$x_{f,t}^j(\tau+1)=y_{f,t}^j(\tau+1)=1\quad\forall f\in F,\forall j\in P_f,CTD_f^j(\tau)<t_\sigma,t=CTD_f^j(\tau), \tag{19}$$

$$x^j_{f,t}(\tau+1) = y^j_{f,t}(\tau+1) = 0 \quad \forall f \in F, \forall j \in P_f, CTD^j_f(\tau) < t_\sigma, t = CTD^j_f(\tau) - 1, \tag{20}$$

$$\begin{aligned} x^j_{f,t}(\tau+1) = y^j_{f,t}(\tau+1) = 1 \quad \forall f \in F, \forall j \in P_f, \forall i \in [1, n_f - 1], P(f,i) = j, \\ P(f,i+1) = j', CTD^j_f(\tau) \geq t_\sigma, CTD^{j'}_f(\tau) < t_\sigma, t = CTD^j_f(\tau), \end{aligned} \tag{21}$$

$$\begin{aligned} x^j_{f,t}(\tau+1) = y^j_{f,t}(\tau+1) = 0 \quad \forall f \in F, \forall j \in P_f, \forall i \in [1, n_f - 1], P(f,i) = j, \\ P(f,i+1) = j', CTD^j_f(\tau) \geq t_\sigma, CTD^{j'}_f(\tau) < t_\sigma, t = CTD^j_f(\tau) - 1. \end{aligned} \tag{22}$$

Constraints (19) and (20) enforce that values, prior to time t_σ, of the decision variables ($x^j_{f,t}(\tau')$ and $y^j_{f,t}(\tau')$) derived from the first round of optimization should be assigned to those new decision variables ($x^j_{f,t}(\tau'+1)$ and $y^j_{f,t}(\tau'+1)$) defined in the same domains (f, j and t). t_1 means the initial time of T, while t_σ represents the initial time defined in the $(\tau+1)$th time period $T(\tau'+1)$.

Constraints (21) and (22) stipulate that for specifically the flights in the air at time t_σ, the new decision variables subject to the second round of optimization must start from the next position after finishing their current flight segment linked by (j, j'). This is because the remaining distance within the segment might be not long enough to realize the amount of LH previously provided by airlines, which, however, is based on the calculation by en entire segment. $t^j_f(\tau')$ and $t^{j'}_f(\tau')$ are the last assigned departure times for flight f.

IV. Illustrative examples

An illustrative example of the methodology introduced in this paper is presented in this section. As stated in [15], network formulations present significant challenges in computational tractability, while some studies have focused exclusively on addressing the computational challenges of the network problem (see [26] for instance). Yet, given the fact that the main objective of this paper is to primarily reveal the effects of including the linear holding practice in ATFM delay assignment, more than improving the model computational performance, we have taken firstly a small sample from the experiments to illustrate the collaborative delay handling process introduced in Sec. III. Some additional experimental materials to assess the computational complexity of the problem are

provided later. GAMS has been used as the modeling tool and Xpress v23.01 optimizer bundled into the GAMS suite has been used as the solver.

A. Case of study setup

As shown in Fig. 8, the data sample chosen for this illustrative example involve 156 flights (red lines) heading towards EHAM airport (yellow label) traversing 1121 elementary sectors (green polygons), with both Estimated Times of Arrival and Estimated Times of Off Block scheduled within the period from 6 AM to 12 AM on October 24, 2016. Initial flight schedules and elementary sector crossings have been taken from the DDR2.

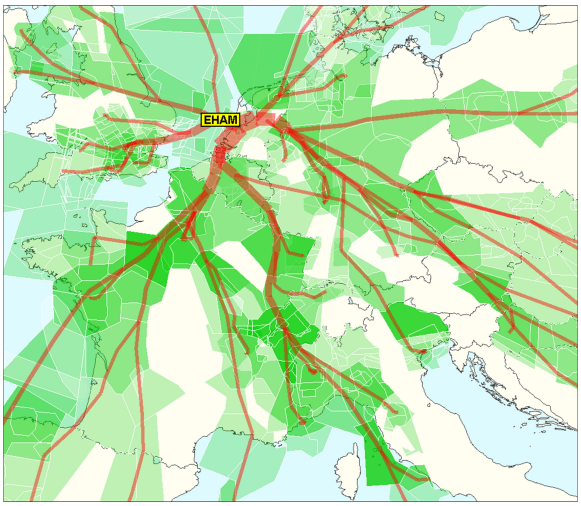


Fig. 8 Flights and associated sectors used in the computational experiments.

Besides the trajectory information, the capacity values of relevant airports and sectors in the case study have been also derived from the DDR2. Airports' operational capacity can be extracted directly from the database, whilst assuming they are equal for departures and arrivals. Obtaining sectors' capacity requires some extra process, and a set of standard source files needed in that database are listed as follows: 1) OpeningScheme.cos, 2) Configuration.cfg, 3) Airspace.spc, 4) TrafficVolume.ntfv, 5) Activation.nact, and 6) Capacity.ncap, with respect to the same Eurocontrol AIRAC (Aeronautical Information Regulation And Control) date.

Since the sectors considered in this paper to define the control "positions" are elementary sectors, which could be combined in realistic operations with other elementary sector(s) and become a collapsed sector during different time period. Therefore, it is needed to obtain a table capturing,

for each time period, the traversed sector's detailed form (i.e., elementary sector itself or collapsed sector it constructs) and the particular form's associated capacity in that time. Specifically, if it is a collapsed sector that contains several elementary sectors, then only the first entered elementary sector (where a control position is defined) for the flight will be counted as one traffic demand of that collapsed sector, and the remaining entires (in the same collapsed sector) will be regarded as internal movement. A general procedure of obtaining this information about sectors' opening scheme and associated capacity is attached in the Appendix, and some examples can be seen in Table 3.

Table 3 Examples of sector opening scheme and capacity values.

Ele. sector	Ini. time	End time	Sector conf.	Capacity	Ini. time	End time	Sector conf.	Capacity	Ini. time	End time	Sector conf.	Capacity
LFMMLE	10:00	10:20	LFMMMALY	38	10:20	10:40	LFMMLYO	44	10:40	11:00	LFMMLYO	44
LFEEYR	10:00	10:20	LFEEHYR	38	10:20	10:40	LFEEHYR	38	10:40	11:00	LFEE5R	42
LFMMLS	10:00	10:20	LFMMMALY	38	10:20	10:40	LFMMLYO	44	10:40	11:00	LFMMLYO	44

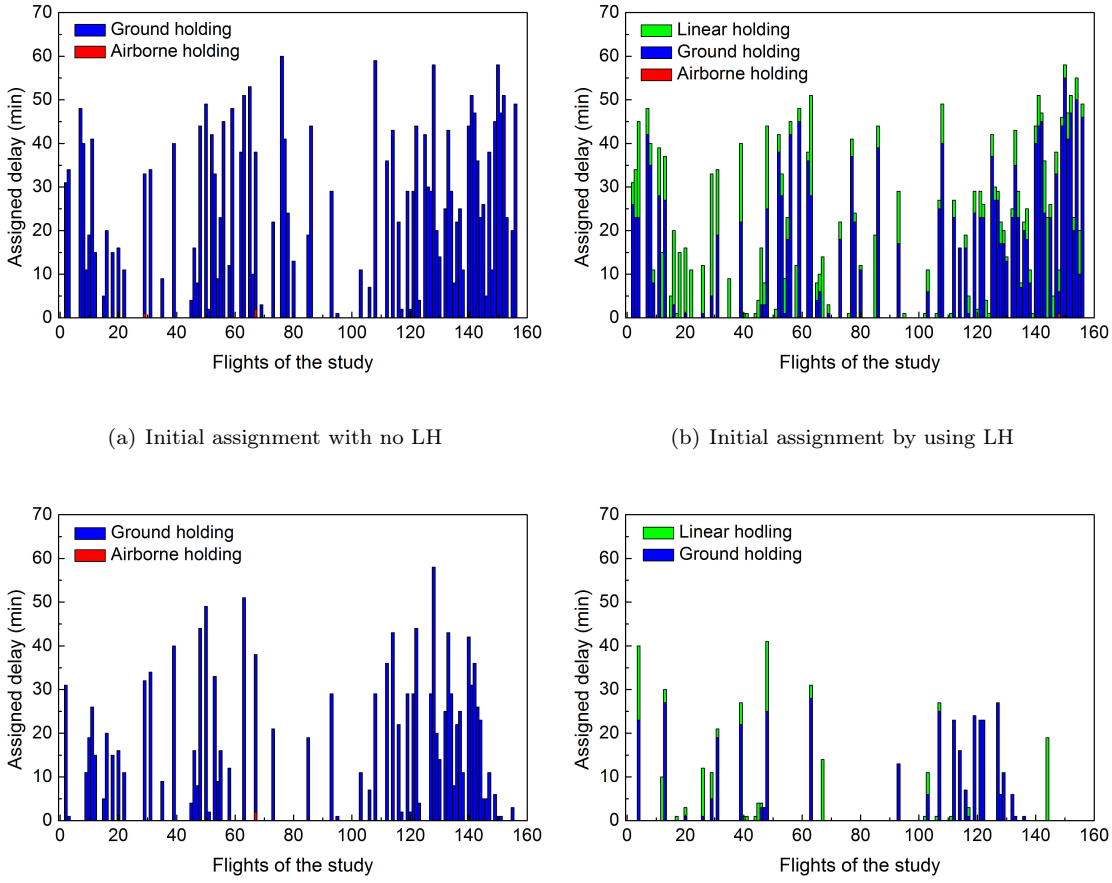
For the initial delay assignment, we assume four hot spots: EHAM airport and sectors EDDDALL1, LFEEKHRZIU and LFEEEUXE; where the demand exceeds the capacity during the studied period. Furthermore, we have also considered a situation where an early capacity recovery occurs at 9 AM, well before scheduled (12 AM), for the above four hot spots, which leads to an update of the delay assignment. It is assumed that updating can be initiated at once while flights can receive and immediately executed the latest delay assignment.

Some other key assumptions have been taken in the computation: 1) the discrete time interval is set to 1 min; 2) $\epsilon = 0.05$ is selected as the fairness factor; 3) the cost weights for airborne holding and linear holding are, respectively, $\alpha = 1.2$ and $\beta = 0.8$ with regards to the ground holding; 4) the LH time bound is approximated as 20% of the planned total trip time based on the statistical average value derived from previous work [7], if not otherwise specified, and are all shared by airlines to the Network Manager; and 5) there are no separation violations or constraints due to other aircraft in the traffic flow.

B. Results of the delay assignment

Figures 9(a) and 9(b) show how in the initial process of delay assignment (i.e., results generated in the first round of model execution), part of the ground holding (and airborne holding) is replaced

by LH. Referring to Table 4, we can see the total delay has a reduction of 120 minutes after this replacement. This is because, including LH means that more space and periods can be used to absorb delays, rather than only at the departure airports prior to take-off. As a consequence, if multiple node constraints occur at the same time, separating delays at different places and periods would contribute to reducing the minimum delay required from multiple constraints. Moreover, we can also notice that more flights are included to share the reduced total delay, leading to an even lower average delay for each flight (see Table 4).



(a) Initial assignment with no LH

(b) Initial assignment by using LH

(c) Updated assignment with no LH

(d) Updated assignment by using LH

Fig. 9 Amount of delay assignment in form of ground holding, airborne holding and linear holding with regards to the four cases of study.

Figures 9(c) and 9(d) illustrate the case when the early capacity recovery occurs (i.e., results yielded in the second round of model execution), assuming that the new round of ATFM delay assignment starts immediately after this recovery. Obviously, the flights that have not been serving

the ground holding, or have been holding on the ground partway, can request for an immediate departure, and thus, have their delays (partially) recovered, as revealed by results of Fig. 9(c).

Table 4 Summarized results for the four cases of study.

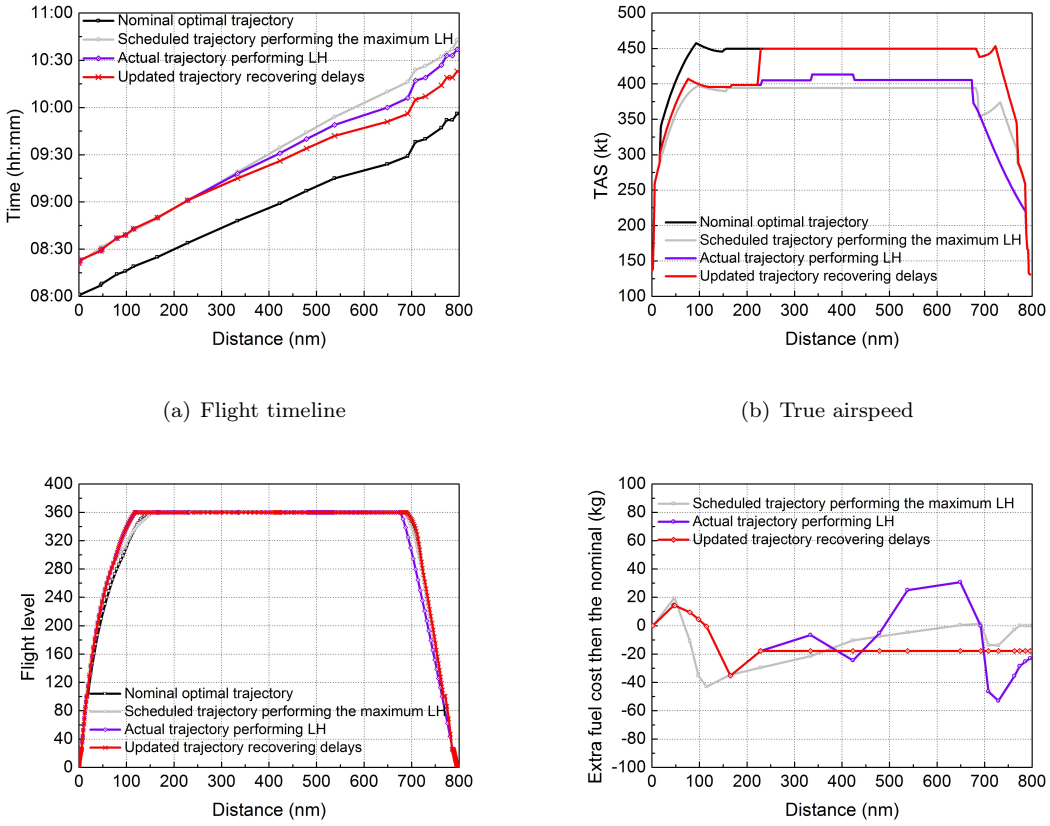
Cases of study	Total delay (min)	Delayed flights (a/c)	Av. total delay (min)	Total AH (min)	Total GH (min)	GH flights (a/c)	Av. GH (min)	Total LH (min)	LH flights (a/c)	Av.LH (min)
Initial delay assignment with no LH	2421	86	28.15	3	2418	86	28.12	0	0	-
Initial delay assignment by using LH	2301	97	23.72	1	1681	72	23.35	619	96	6.45
Updated delay assignment with no LH	1369	66	20.74	2	1367	66	20.71	0	0	-
Updated delay assignment by using LH	499	38	13.13	0	370	27	13.70	129	24	5.38

When implementing LH, however, the remaining total delay reduces remarkably once the delay assignment is updated (see Fig. 9(d)). There are two main reasons that could account for these promising results. First, benefiting from the shortening of ground holding, the departure time of one flight can be advanced. Once the delay is updated, less ground holding, and thus less total delay will be realized, as exactly is the case shown in Fig. 9(c). Since most of the flights are observed to substitute part of the ground holding by LH (see Table 4), the effects can be enlarged notably. The second reason is because the flexibility of LH compared to the ground holding in terms of delay absorption, as mentioned previously in Sec. II A. Regarding this scenario, a detailed analysis is given in the next section.

C. Delay recovery for an airborne flight performing LH

In this section, the same flight (LIRF-EHAM) introduced in Sec. III B is analyzed in detail. During the initial process of delay assignment, this particular flight is allocated with 41 min of delay in total (imposed on the arrival slot), while allowed to wait on the ground for 22 min (i.e., ground holding) but flying slower to absorb the rest of the delay, i.e., 19 min, by means of LH in the air (assuming no update occurs).

After serving 22 min of ground holding followed by encountering the update of delay assignment, at 9 AM, as shown in Fig. 10(a), the flight starts to recover its nominal trajectory. The process is initiated when passing the next designated position (229 nm) compared to the flight's current geographical position. Afterwards, the updated timeline (the red line) deviates from the actual one performing an amount of LH (19 min, the blue line) which is lower than its maximum LH (25 min, the grey line) shared to the NM. It can be noticed that the slope of the red curve becomes flatter



(a) Flight timeline

(b) True airspeed

(c) Vertical trajectory

(d) Extra fuel consumption

Fig. 10 Effects of delay recovery for the flight (LIRF-EHAM) performing LH partway in the air when encountering the update of delay assignment.

as to be exactly parallel to the nominal timeline (the black line) during the remaining distance. At the end, there are 14 min of delay saved, reducing the total delay from 41 min to 27 min.

As we can see from Fig. 10(b), the TAS of the actual trajectory performing 19 min of LH lies between the nominal TAS and the one having the maximum LH. Interestingly, since delays are not assigned evenly along the trajectory, we may notice that the actual speed (the blue line) changes progressively during the cruise phase (due to the discrete time step of 1 min in this paper), which may result in an increase in flight crew workload. However, given the Required Time of Arrival (RTA) featured in modern on-board FMS, and aimed at autopilot when performing LH, it might not raise too much concerns on the procedures. As for the vertical trajectory illustrated in Fig. 10(c), caused from the changes of climb and descent speeds seen in Fig. 10(b), the geographical positions of TOC and TOD vary from the nominal trajectory.

At last, as shown in Fig. 10(d), recall again that the difference of LH with respect to typical airborne holding is located at whether the extra fuel needs to be consumed. Obviously, without this premise, LH might not be favored by airlines, given a safer and cheaper ground holding is always there. As a consequence, by restricting the fuel along the whole trajectory when optimizing it for LH, the fuel consumption can be constrained to the nominal one. Note, however, that due to limitations of the model (fuel constraint is enforced for each discrete flight segment), the fuel consumption will not be exactly the same as the nominal one and small differences can be appreciated. Consequently, as the trajectory updating (red line) occurs at the position where less fuel has been burnt (i.e., 229 nm) than the nominal, and keeps the same unit fuel consumed (or specific range, because the initially scheduled airspeed is recovered) for the remaining distance, such that the final difference in fuel is exactly the same as that observed at the distance of 229 nm.

D. Extension of the case study: sensitivity analysis and scalability test

With the aim to demonstrate how some of the key algorithm parameters affect ground holding and airborne holding as well as linear holding, more computational experiments have been conducted, with a sensitivity analysis presented in this section. Table 5 shows the relevant independent parameters and their associated ranges considered for the design tradeoffs. In addition, tests of computational scalability have been also performed under three typical ATFM scenarios given different problem sizes. These numerical experiments have been run on a 64 bit Intel i7-4790 @ 3.60 GHz quad core CPU computer with 16 GB of RAM memory and Linux OS. GAMS v.24.02 software suite has been used as the modeling tool and Gurobi v.7.02 optimizer has been used as the solver.

Table 5 Values of independent parameters in the sensitivity study.

Parameters	Baseline	Min	Max	Step	Comment
ε	0.05	0	0.5	0.05	fairness factor of delay assignment
α	1.2	1.2	3	0.2	cost weight of airborne holding to ground holding
β	0.8	0.1	1.1	0.1	cost weight of linear holding holding to ground holding
τ	4	1	7	1	time period when capacity updated
$C^{\%}(\tau)$	120%	80%	200%	10%	updated capacity with regard to the initial

A benchmark scenario taken for the sensitivity analysis involves 1131 flights traversing across 164 elementary sectors in the French airspace between 10 AM to 12 AM, July 28, 2016, as shown in Fig. 11. The unit period of capacity is set to per 20 min for all the (active) collapsed sectors

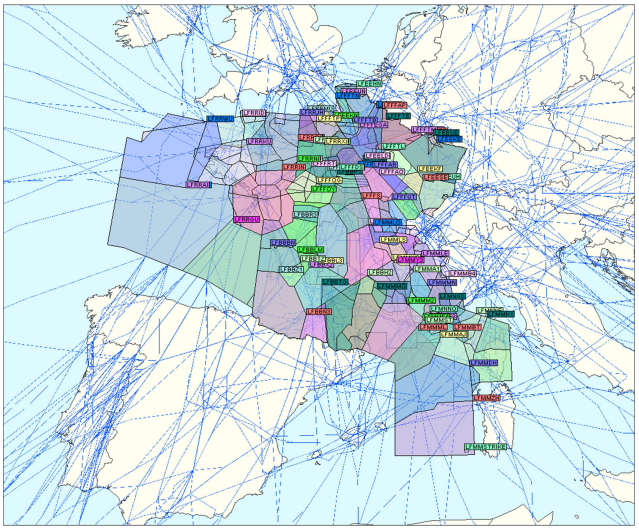


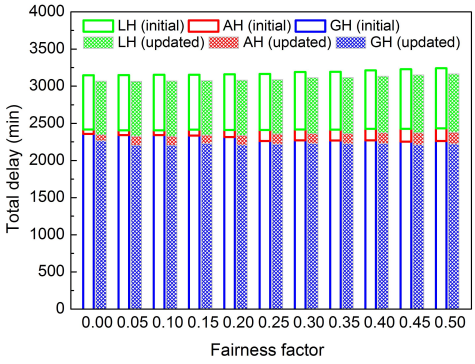
Fig. 11 Flights and associated sectors used in the computational experiments.

(123 in total in the case). As mentioned in Sec. IV A, the elementary sectors might be combined to different collapsed sectors during each time period. Initial trajectory information and capacity data are processed in the same way as done for the previous case study. Parameters' baseline values are as shown in Table 5, and the sensitivity experiments are conducted by varying each individual parameter within its predefined range whilst fixing the rest by their baseline values.

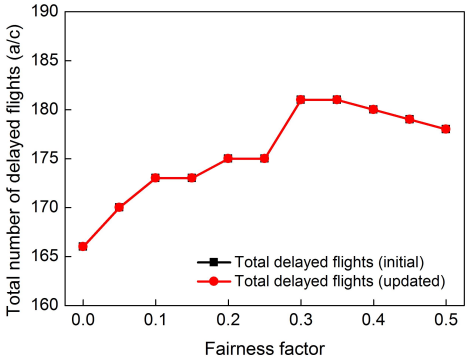
Changes of parameters ϵ, α and β would affect both the initial delay assignment and its subsequent (potential) update. Results of their sensitivity study are presented in Fig. 12. The amount of imposed delays and the number of affected flights are considered, for each parameter, with respect to four indicators including total delay and individual delay realized by ground holding, airborne holding and linear holding respectively.

The effects of increasing fairness factor (ϵ) can be appreciated from Figs. 12(a) and 12(b). Given an improved equality in the delay assignment process, the total delay that is required as a whole grows gradually, and it is the increased amount of airborne and linear holding that contributes to this growth as ground holding on the contrary decreases slightly. On the other hand, more flights are involved in the regulation as expected to share the higher amount of delay.

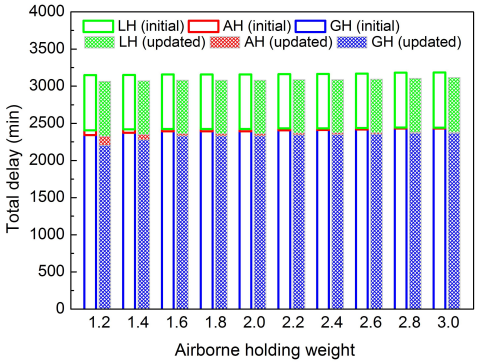
Figs. 12(c) and 12(d) show the cases when varying the weighted cost of airborne holding. It is worth noting that the updated amount of airborne holding turns even higher than that resulting from the initial delay assignment when the parameter's value is relatively low (e.g., $\alpha = 1.2$). This



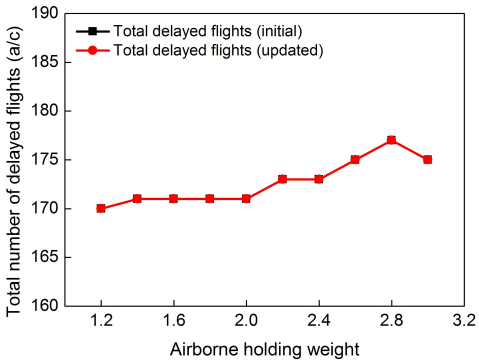
(a) Fairness factor - Amount of delay



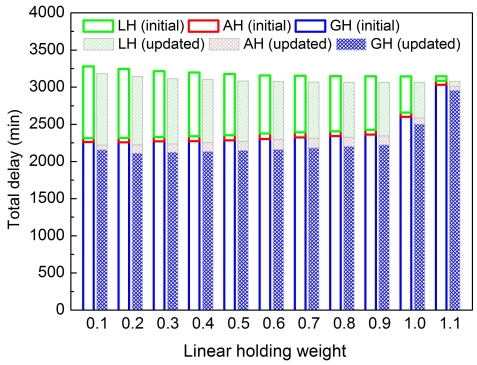
(b) Fairness factor - Number of affected flights



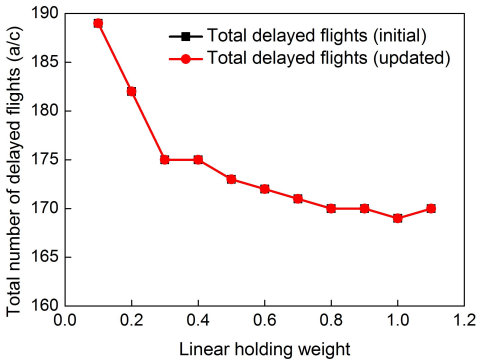
(c) Airborne holding cost - Amount of delay



(d) Airborne holding cost - Number of affected flights



(e) Linear holding cost - Amount of delay

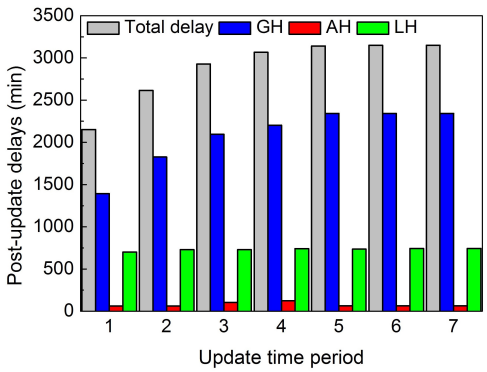


(f) Linear holding cost - Number of affected flights

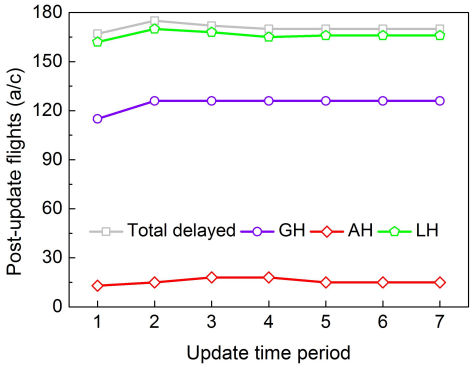
Fig. 12 Sensitivity analysis on key parameters of model formulation with respect to the amount of delay assignment and the number of affected flights for each holding practice.

is due to the fact that if performing airborne holding is only slightly more expensive than ground holding (and linear holding), then it could be better to impose delays directly at certain sectors of reduced capacity (by means of air holding), rather than transferring the overall delay through

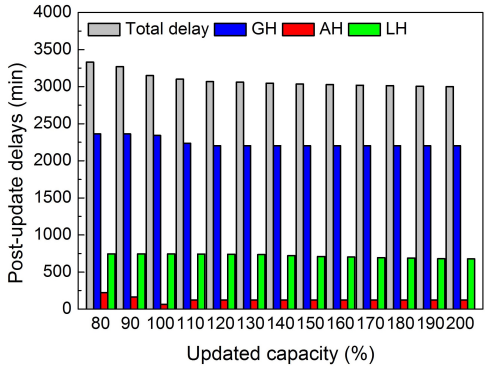
multiple sectors until the origin airports to execute ground holding which though is cheaper.



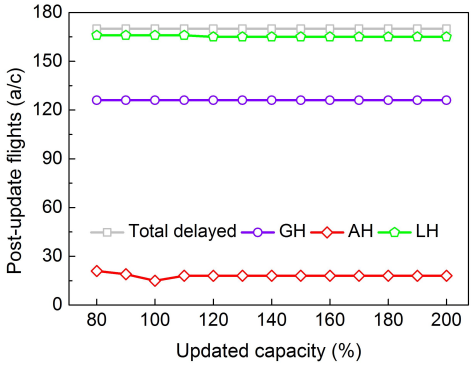
(a) Update time - updated delay assignment



(b) Update time - updated affected flights



(c) Update capacity - updated delay assignment



(d) Update capacity - updated affected flights

Fig. 13 Sensitivity analysis on capacity updating parameters with respect to the post-update delay assignment and the number of affected flights for each holding practice.

The sensitivity results for linear holding's weighted cost are as shown in Figs. 12(e) and 12(f), from which it can be noticed that when its value is greater than 1 (i.e., higher than the cost of ground holding), the assigned delays decrease remarkably while the reduced part is instead realized by ground holding. However, although the total delay is almost unchanged, the number of delayed flights reduces significantly, meaning that the average delay for those affected flights would increase as well. In other words, with greater preference for linear holding, more flights would be captured in the regulation to share a constant total delay.

Meanwhile, the parameters τ and $C^s(\tau)$, other than ϵ, α and β, would only affect the delay updating process, and their results are summarized in Fig. 13. Generally, the earlier the capacities

start to recover and the higher the updated capacities become, the less delays are required in the updating process. However, after the 5th time period (i.e., 11:20 AM) the delays that can be saved are almost constant, which is due to the fact that most of the flights have departed before that time, and therefore the already realized ground holding, prior to takeoff, cannot be recovered even with the increased capacities.

Moreover, in order to understand what problem size can be handled, by executing the model of this paper, within a reasonable time given current optimization tools, two additional scenarios have been further considered (in addition to the benchmark scenario of the sensitivity experiment, which is labeled as S1) in the scalability tests. Namely, 2 hours' traffic across the ECAC (European Civil Aviation Conference) area, labeled as S2, and 24 hours' traffic across the French airspace, labeled as S3 were considered. Scenario configurations are summarized in Table 6, including the number of flights, time periods, elementary and collapsed sectors for instance. Tests of the initial delay assignment and the subsequent updating are conducted respectively for each scenario. Finally, the discrete time interval and the unit period of capacity are still set to 1 min and 20 min respectively, while the maximum delay allowed to each flight (i.e., solution search space) is limited to 180 min for the sake of reducing the total number of decision variables.

Table 6 Scenario setup for the scalability tests.

Scenarios	Duration	Scope	Flights	Ele. Sectors	Col. Sectors	Time periods	Upd. time
S1	2 hour	France	1131	164	123	6	4
S2	2 hour	ECAC	3835	1626	1059	6	4
S3	24 hour	France	6089	164	205	72	37

Table 7 Problem size and computational time for each test of the study.

Scenarios	Variables	Constraints	Non-zero elements	Generation time (s)	Comput. time (s)
S1_initial	575,101	1,193,063	3,605,217	308	238
S1_update	569,965	1,183,343	3,577,328	311	94
S2_initial	2,623,501	5,507,530	13,648,395	2184	665
S2_update	2,611,629	5,487,453	13,586,808	2168	34
S3_initial	2,743,201	5,825,900	14,483,261	1723	2566
S3_update	2,733,569	5,809,513	14,432,863	1519	171

The size of the problem and the computational time, for each scenario taken into the model, are

Table 8 Results of delay assignment realized by the three holding practices.

Scenarios	Ground holding		Airborne holding		Linear holding		Total delay	
	Time (min)	Flights (a/c)	Time (min)	Flights (a/c)	Time (min)	Flights (a/c)	Time (min)	Flights (a/c)
S1_initial	5155	308	338	39	898	268	6391	384
S1_update	4353	286	240	31	878	276	5471	361
S2_initial	13192	655	168	25	2063	400	15423	791
S2_update	12237	613	126	18	1929	376	14292	738
S3_initial	31456	1044	412	25	2440	835	34308	1157
S3_update	8604	650	45	10	1647	570	10296	796

summarized in Table 7, while detailed results of the delay assignment can be found in Table 8. Note that the generation time shown in the table includes the compilation time (for reading input files), execution time (for numerical calculations on existing data) and generation time (for constructing the equations and calling the solver), and this particular time could be saved remarkably for a second round of generation where only a few parameters change, such as the delay updating process, by means of using the save and restart option of GAMS. The solver's computational times recorded in Table 7 show that the optimal solution of each scenario could be found within a reasonable time. Nevertheless, as there are numerous factors that may have effects on an MILP model's computational performance, such as the problem size (e.g., variables and constraints), the tightness of the formulation and the heuristic method used in the solver, it is quite difficult to only select a limited number of analytical indicators to predict the corresponding computational time. Empirical studies would be an appropriate way to quantify the actual scale that is tractable for this model, and deserve further study in the future.

V. Conclusions

In this paper, a cost-based LH practice was merged into the ATFM regulation for delay absorption, together with the commonly seen ground and airborne holdings. In the light of trajectory based operations, airlines' sharing of maximum LH bounds derived from their own optimal aircraft trajectory generation, could be effectively utilized by the ATFM side as one of the optimization factors considered for delay assignment.

Incorporating the LH means that more space and periods in the network can be used to absorb delays. Provided multiple node constraints occur at the same time, splitting delays at different

places and times could contribute to reducing the minimum system delay required from multiple constraints. Results suggest that replacing ground holding with LH enables aircraft to depart earlier, with less ground holding. This provides more flexibility in responding to changes in capacity, and thus less total delay would be realized. Moreover, if the delays are canceled ahead of schedule, aircraft already airborne and performing LH, could accelerate to the speed as initially planned and recover part of the delay at no extra fuel cost.

Nevertheless, it must be noted that in actual operations not all aircraft flying along a route in a sector are in line going to the same direction, such that a variety of aircraft performing LH (with speed reduced) would invariably induce conflicts with the traffic flow along side. This paper accounts for only the sector's entrance rate as a (capacity) constraint and thereby impose a controlled time at each sector's entry position, while neglecting the specific movements inside the sector (and the airport). In other words, the results of this paper represent best-case outcomes assuming no traffic conflicts or other constraints specified by detailed sector and airport (e.g., the runway usage) operations.

Future work will focus on extending airlines' participation in the ATFM process to further include re-routing and flight level changing options, which correspond to the aircraft lateral and vertical (alternative) trajectory planning respectively, in addition to this paper's delay absorption method realized by the three holding practices. Choosing an alternative trajectory might help to avoid the traffic hotspots, and thus to reduce the flight's potential delay costs that will be incurred subsequently, but each alternative trajectory itself represents a different operating cost determined by route distance, airspace charges and fuel consumption. For those airlines willing to take part in this collaborative decision-making process, more information (e.g., alternative trajectory with its preference) will be required to share with the Network Manager, and a centralized optimization model, on the other hand, will be needed to integrate these options.

Furthermore, given that the ATFM regulations are typically issued in response to severe weather conditions, the wind field and non-standard atmospheres, which always have great effects on real flights, should be further considered. This, however, could be processed from the airline's side as well, and eventually result in some trajectory associated information which, in turn, could be taken

into account by the Network Manager. Finally, as the problem becomes larger and more complex, empirical studies on the model's tractability are also required for the future work.

Acknowledgment

The authors would like to thank Airbus Industrie for the use of PEP (Performance Engineers Program) suite, which allowed us to undertake realistic aircraft performance parameters to conduct numerical simulations of trajectory optimization. This research is partially supported by grants from the Funds of China Scholarship Council (201506830050) and by a La Caixa-UPC International Mobility Scholarship (Y4262159C). The authors would also like to thank the associate editor and the reviewers for their insightful comments that have helped us to improve this paper.

[1] FAA, *Collaborative Trajectory Options Program (CTOP): Document Information. Federal Aviation Administration, AC 90-115.*, 2014.

[2] Miller, M. E. and Hall, W. D., "Collaborative Trajectory Option Program demonstration," *34th IEEE/AIAA Digital Avionics Systems Conference (DASC), Prague, Czech Republic*, IEEE, 2015, pp. 1C1–8.

[3] Ball, M. O., Hoffman, R., Lovell, D., and Mukherjee, A., "Response mechanisms for dynamic air traffic flow management," *Proceedings of the 6th Europe-USA ATM Seminar. Baltimore. US*, 2005.

[4] Prats, X. and Hansen, M., "Green Delay Programs: absorbing ATFM delay by flying at minimum fuel speed," *Proceedings of the 9th USA/Europe air traffic management R&D seminar, Berlin, Germany*, 2011.

[5] Delgado, L. and Prats, X., "En route speed reduction concept for absorbing air traffic flow management delays," *Journal of Aircraft*, Vol. 49, No. 1, 2012, pp. 214–224.

[6] Delgado, L. and Prats, X., "Operating cost based cruise speed reduction for ground delay programs: Effect of scope length," *Transportation Research Part C: Emerging Technologies*, Vol. 48, 2014, pp. 437–452.

[7] Xu, Y., Dalmau, R., and Prats, X., "Maximizing airborne delay at no extra fuel cost by means of linear holding," *Transportation Research Part C: Emerging Technologies*, Vol. 81, 2017, pp. 137–152.

[8] Jones, J. C., Lovell, D. J., and Ball, M. O., "En route speed control methods for transferring terminal delay," *Proceedings of the 10th USA/Europe Air Traffic Management Research and Development Seminar, Chicago, USA*, 2013.

[9] Günther, T. and Fricke, H., "Potential of speed control on flight efficiency," *Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT), Belgrade, Serbia*, Vol. 1, 2006, pp. 197–201.

[10] Airservices, *Airservices Australia Annual Report 2006-2007, Canberra, Australia*, 2007.

[11] Odoni, A. R., "The flow management problem in air traffic control," *Flow control of congested networks*, Springer, 1987, pp. 269–288.

[12] Terrab, M. and Paulose, S., "Dynamic strategic and tactical air traffic flow control," *IEEE International Conference on Systems, Man and Cybernetics, Chicago, USA*, IEEE, 1992, pp. 243–248.

[13] Bertsimas, D. and Patterson, S. S., "The air traffic flow management problem with enroute capacities," *Operations research*, Vol. 46, No. 3, 1998, pp. 406–422.

[14] Lulli, G. and Odoni, A., "The European air traffic flow management problem," *Transportation Science*, Vol. 41, No. 4, 2007, pp. 431–443.

[15] Bertsimas, D. and Gupta, S., "Fairness and collaboration in network air traffic flow management: an optimization approach," *Transportation Science*, Vol. 50, No. 1, 2015, pp. 57–76.

[16] Samà, M., D'Ariano, A., D'Ariano, P., and Pacciarelli, D., "Optimal aircraft scheduling and routing at a terminal control area during disturbances," *Transportation Research Part C: Emerging Technologies*, Vol. 47, 2014, pp. 61–85.

[17] Samà, M., D'Ariano, A., D'Ariano, P., and Pacciarelli, D., "Air traffic optimization models for aircraft delay and travel time minimization in terminal control areas," *Public Transport*, Vol. 7, No. 3, 2015, pp. 321–337.

[18] Samà, M., D'Ariano, A., Pacciarelli, D., Palagachev, K., and Gerdts, M., "Optimal aircraft scheduling and flight trajectory in terminal control areas," *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy*, IEEE, 2017, pp. 285–290.

[19] Samà, M., D'Ariano, A., Corman, F., and Pacciarelli, D., "Coordination of scheduling decisions in the management of airport airspace and taxiway operations," *Transportation Research Procedia*, Vol. 23, 2017, pp. 246–262.

[20] FAA, *Instrument Procedures Handbook, FAA-H-8083-16A, Chapter 3, Arrivals*, 2015.

[21] Airbus, *Flight Crew Operation Manual (FCOM): A320, Version 1.3.1*, 1993.

[22] Jones, J. C., Lovell, D. J., and Ball, M. O., "Combining Control by CTA and Dynamic En Route Speed Adjustment to Improve Ground Delay Program Performance," *Proceedings of the 11th USA/Europe Air Traffic Management Research and Development Seminar, Lisbon, Portugal*, 2015.

[23] Dalmau, R. and Prats, X., "Fuel and time savings by flying continuous cruise climbs: Estimating the benefit pools for maximum range operations," *Transportation Research Part D: Transport and Environment*, Vol. 35, 2015, pp. 62–71.

[24] Bolić, T., Castelli, L., Corolli, L., and Rigonat, D., "Reducing ATFM delays through strategic flight planning," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 98, 2017, pp. 42–59.

[25] Mukherjee, A. and Hansen, M., "A dynamic stochastic model for the single airport ground holding problem," *Transportation Science*, Vol. 41, No. 4, 2007, pp. 444–456.

[26] Rios, J. and Ross, K., "Massively parallel dantzig-wolfe decomposition applied to traffic flow scheduling," *Journal of Aerospace Computing, Information, and Communication*, Vol. 7, No. 1, 2010, pp. 32–45.

Appendix

Procedure of generating the operating capacities of active sectors in different time periods.

def Generation-of-Active-Sector-and-Capacity**:**

open 'OpeningScheme.cos' and 'Configuration.cfg'**:**

Create a dictionary of each ACC's opening scheme mapping to its associated elementary and/or collapsed sectors

open 'Airspace.spc'**:**

Create a dictionary of each collapsed sector's belonging elementary sectors

for elementary sector **in** list-of-designed-positions**:**

 # the list contains all the elementary sectors that the scheduled trajectories traverse

for period **in** list-of-time-periods:

 # the list contains all the time periods taken into the case study

Cross reference between the ACC's dictionary and collapsed sector's dictionary

Fill in a data frame (frame-of-elementary-sector)

 # the frame shows for each elementary sector what the operating sector (being itself or grouped into some collapsed sector) it will be during each period of time

open 'TrafficVolume.ntfv' and 'Activation.nact'**:**

Create a dictionary of each elementary/collapsed sector's associated Traffic Volume(s) (TV)

Remove those not activated *# with only one activated TV remained for each sector*

open 'Capacity.ncap'**:**

Create a dictionary of each unit's capacity and capacity-active period(s)

for item, time **in** frame-of-elementary-sector**:**

if item **in** list-of-units **and** time **in** list-of-unit-active-times**:**

 # search in the above capacity-dictionary

item**.append**(unit's capacity)

elif item **in** list-of-traffic-volumes**:** *# i.e., traffic-volume-dictionary.keys()*

tv = traffic-volume-dictionary[item][i] *# i.e., name of the traffic volume*

if tv **in** list-of-units **and** time **in** list-of-unit-active-times**:**

item**.append**(tv's capacity)

return Data-Frame-of-Active-Sector-and-Capacity
